

ABSTRACT OF THE DISCLOSURE

A method factors an input finite state transducer (FST) with unknown symbols into a left-sequential FST and a right-sequential FST while avoiding direct factorization of the unknown symbols. The left-sequential FST is formed by replacing
5 each occurrence of the unknown symbol in the input FST with a sequence of the unknown symbol and a diacritic. The right-sequential FST is formed by replacing each occurrence of the diacritic with a symbol representative of an empty string and an output symbol.